Electronic Markets and Limit Order Book

Business Intelligence per i Servizi Finanziari 2023-2024

Antonio Candelieri

Trading in Electronic Market

- Orders are managed by a **matching engine** and a **limit order book** (**LOB**)
- The LOB keeps track of incoming and outgoing orders
- The matching engine uses a well-defined algorithm that establishes when a possible trade can occur, and if so, which criterion is going to be used to select the orders that will be executed

LOB and (its) spread

- The LOB is defined on a fixed discrete grid of prices (the price levels)
- The size of the step (the difference between one price level and the next) is called the **tick**, and in the US the minimum tick size is 1 cent for all stocks with a price above one dollar
- In other markets several different tick sizes coexist. For example, in the Paris Bourse or the Bolsa de Madrid, tick sizes can range from 0.001 to 0.05 euros depending on the price the stock is trading at
- The difference between the ask and the bid price, **the quoted spread** is

$$Quoted\ Spread_t = P_t^a - P_t^b$$

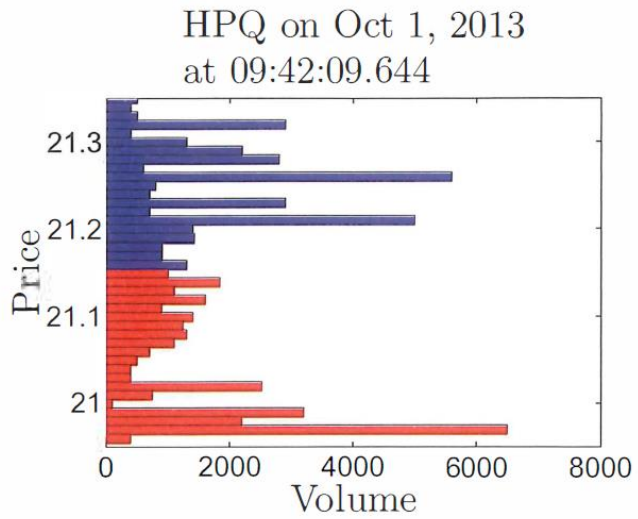
where P_t^a and P_t^b are the best ask and the best bid prices, respectively

LOB and midprice

- Sometimes the bid is equal to the ask and the spread is zero
- In this case, the market becomes locked, but if this happens, it tends not to last long - although for some very liquid assets it is becoming an increasingly more frequent event
- Another common object used when describing the LOB is the **midprice**: the arithmetic average of the bid and the ask.
- Midprice is often used to proxy for the true underlying price of the asset - the price for the asset if there were no explicit or implicit trading costs (and hence no spread)

LOB

- On the left: LOB's HPQ, a frequently traded and liquid asset
- HPQ's LOB has LOs posted at every tick out to (at least) 20 ticks away from the mid price
- On the right: FARO's LOB
- FARO is a seldom traded, illiquid asset. This asset has thinly posted bids and offers and irregular gaps in the LOB



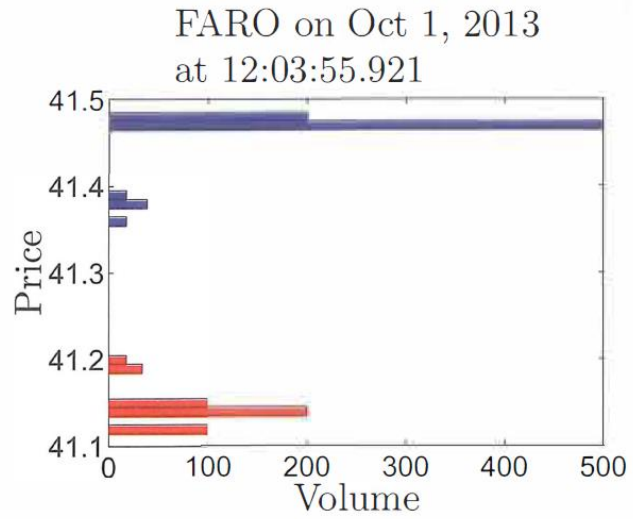


Figure 1.1 Snapshots of the NASDAQ LOB after the 10,000th event of the day. Blue bars represent the available sell LOs, red bars represent the available buy LOs.

Trading in Electronic Market

- Most markets prioritize **MOs** (**Market Orders**) over **LOs** (**Limit Orders**) and then use a price-time priority whereby, if an MO to buy comes in, the buy order will be matched with the standing LOs to sell in the following way:
  - first, the incoming order will be matched with the LOs that offer the best price (this means that for buy orders, the sell LOs with the lowest price),
  - then, if the quantity demanded is less than what is on offer at the best price, the matching algorithm selects the oldest LOs, the ones that were posted earliest, and executes them in order until the quantity of the MO is executed completely.
- If the MO demands more quantity than that offered at the best price, after executing all standing LOs at the best price, the matching algorithm will proceed by executing against the LOs at the second-best price, then the third-best and so on until the whole order is executed.
- LOs that have increasingly worse prices are named LOs that are deeper in the LOB, and the process whereby an entering market order executes against standing LOs deeper in the LOB is called '**walking the book**'

The Limit Order Book

- Addition of LO to LOB. As mentioned above, electronic exchanges are, at their most basic, described by an LOB and a matching algorithm.
- How price-time priority works: an incoming LO joins the LOB at the order's price and is placed last in the execution queue at that price.
- In the figure, LOs are displayed as blocks of length equal to their quantities.
- LOs are ordered in terms of time priority from right to left, so that when a new buy LO comes in at $23.09 (the purple block) it will be added to the line of blocks already resting at that price.
- This new LO joins the queue at the point closest to the y-axis, becoming the third LO waiting to be executed at $23.09.

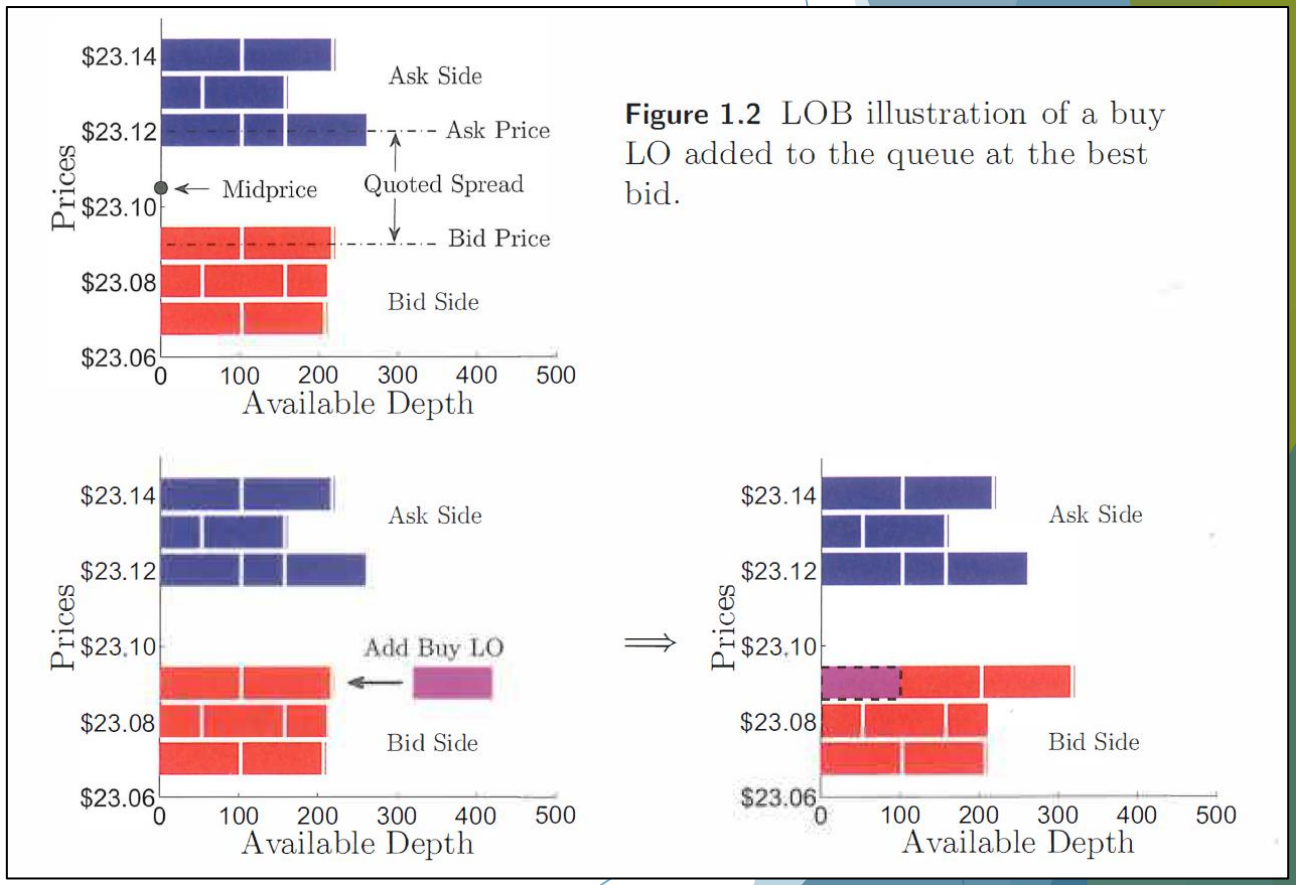


Figure 1.2 LOB illustration of a buy LO added to the queue at the best bid.

The Limit Order Book

- MO walks the LOB or is re-routed
- Suppose we are looking at the venue with the LOB depicted at the top of the first figure
- Assume that this venue's best bid is the best buy quote that the market, across all venues, currently displays
- A new MO (to sell) 250 shares enters this market as depicted by the sum of the green blocks in the top of the second figure.
- The matching engine goes through the LOB, matching existing (posted) LOs (to buy on the bid side) with the entering MO following the rules in the matching algorithm.
- In the LOB there are two LOs at the best bid \$23.09, represented by the two red blocks, both for 100 units, totalling 200 units. These 200 units are executed at the best bid.

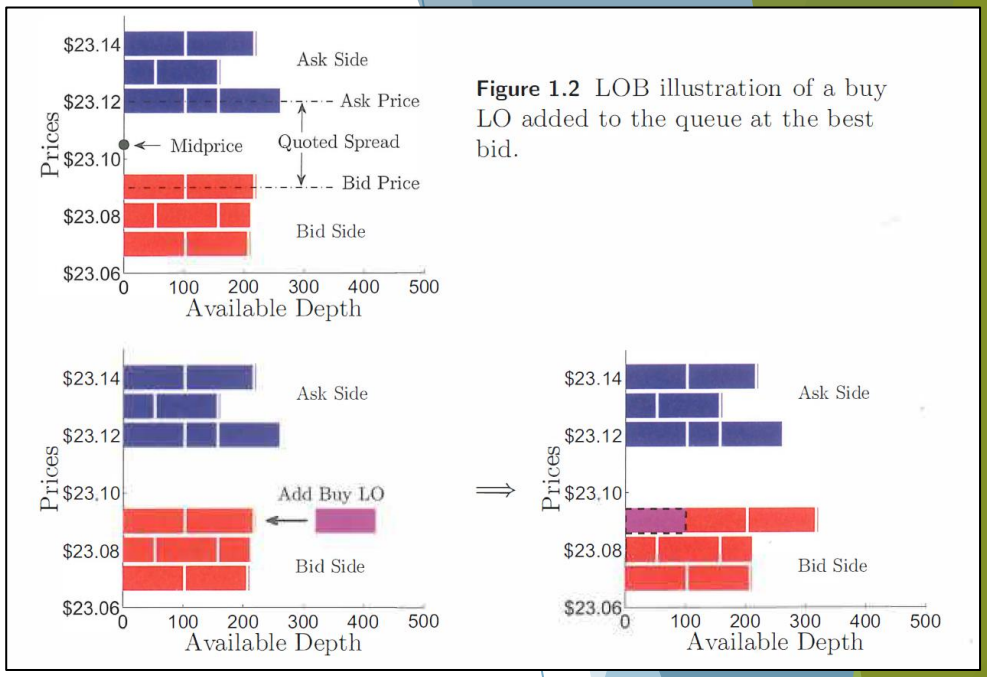


Figure 1.2 LOB illustration of a buy LO added to the queue at the best bid.

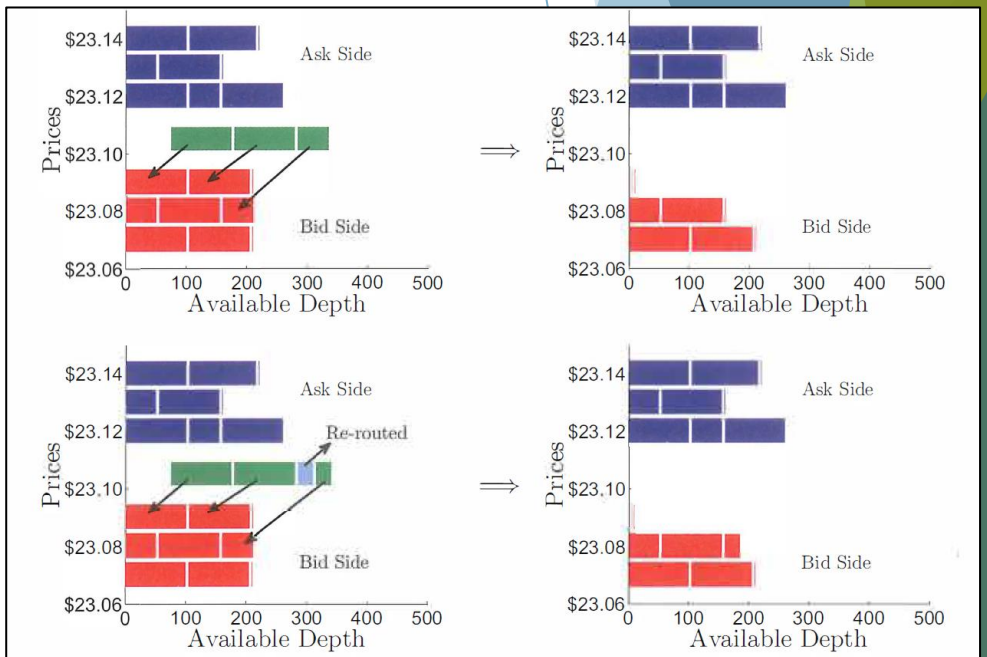


Figure 1.3 LOB illustration of a sell MO walking the LOB with and without

The Limit Order Book

- What happens to the final 50 units depends on the order type and the market it is operating in
- In a standard market, the remaining 50 units will be executed against the LOs standing at $23.08 ordered in terms of time-priority (the MO will 'walk the book')
- This is captured by the top of the figure: the left panel shows that the MO coming in is split into three blocks, the first two are matched with LOs at $23.09 and the last with the LOs at $23.08
- After the MO is fully executed the remaining LOB is shown in the top right panel of the figure

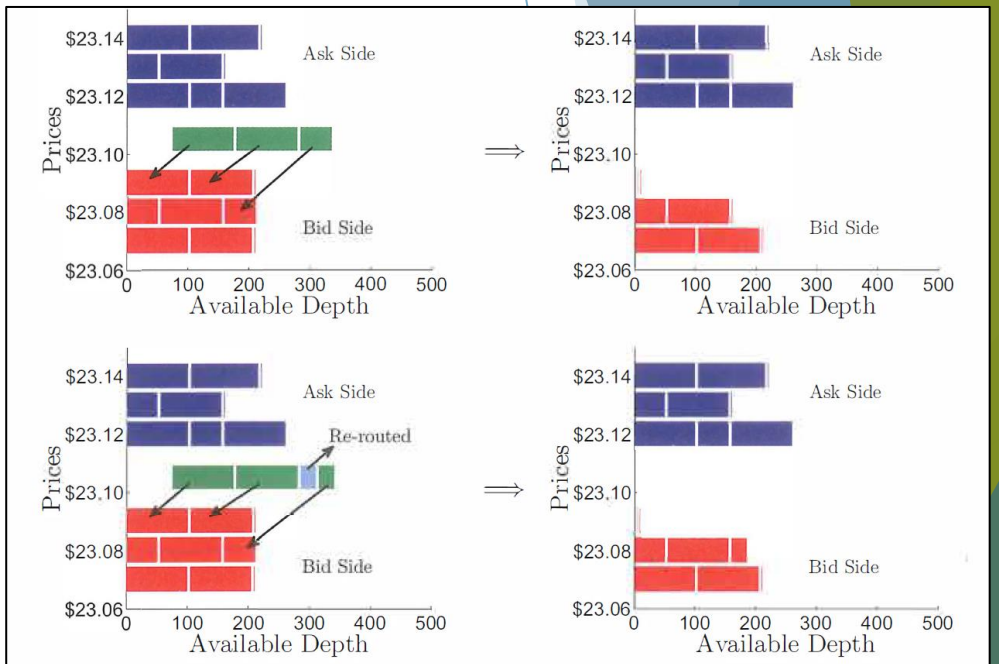


Figure 1.3 LOB illustration of a sell MO walking the LOB with and without

The Limit Order Book

- in the US, there are order protection rules to ensure MOs get the best possible execution, and which (depending on the order type) may require the exchange to re-route the remaining 50 units to another exchange that is also displaying a best bid price of $23.09
- As shown in the bottom left of figure, part of the remaining 50 units (the light blue block) is re-routed to another venue(s) with liquidity posted at $23.09. Only once all liquidity at $23.09 in all exchanges is exhausted, can the remaining shares of the MO return and be executed in this venue against any LO resting at (the worse price of) $23.08. In this example, 25 units were re-routed to alternate exchanges, and 25 units returned to this venue and walked the book
- The MO could be an **Immediate-or-Cancel** (**IOC**) order, specifying that the remaining 50 shares that cannot be executed at the best bid should be cancelled entirely
- Because of these order protection rules (trade-through rules - there is no such rule in European markets), you will very seldom observe in the US an MO walking the book straight away. Rather, you may see a large MO being chopped up and executed sequentially in several markets in a very short span of time. This also implies that as depth disappears an MO at the end of a sequence of other orders may be executed against very poor prices, and, in the worst circumstances it may be matched with stub quotes - LOs at prices so ridiculous that clearly indicate they are not expected to be executed Thus, the LOB serves to keep track of LOs and apply the algorithm that matches incoming orders to existing LOs

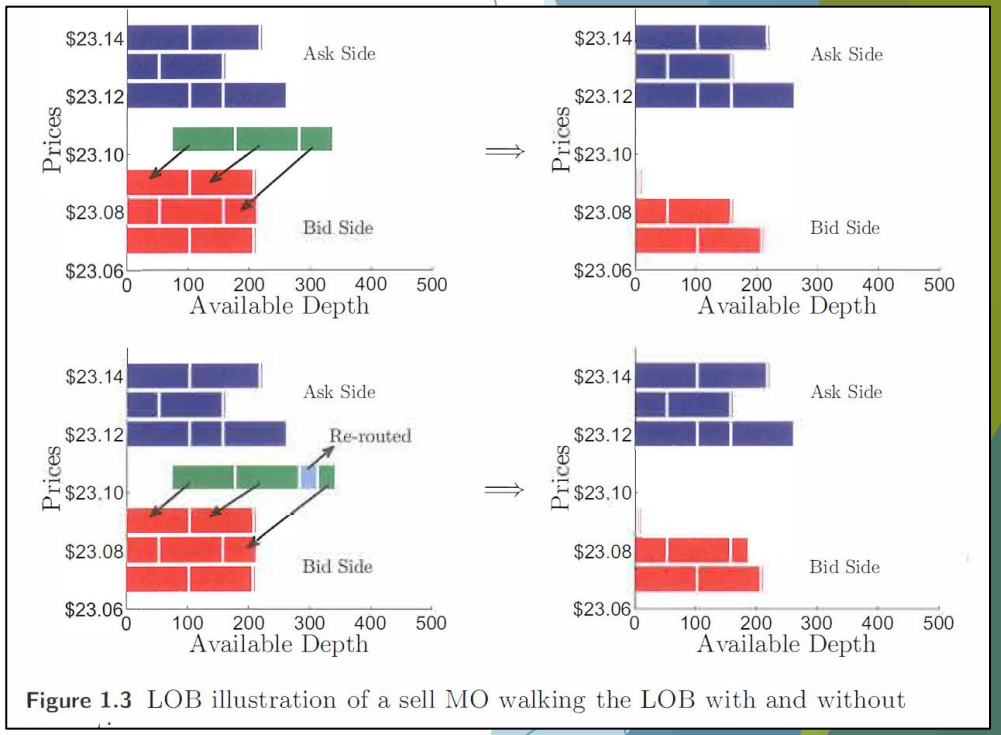


Figure 1.3 LOB illustration of a sell MO walking the LOB with and without

- In the figure, you can see how the LOB evolves through time (over 5 minutes) for three different stocks (i.e. HPQ, NTAP and ORCL)
- On the x-axis is time in minutes, and on the y-axis are prices in dollars
- The blue regions on top represent the ask side of the LOB, the posted sell volume, while the bid side is below in red, showing the posted buy volume
- The best prices, the bid and ask are identified by the edges of the intermediate light shaded beige region, which identifies the bid-ask spread
- Volume at each price level is illustrated by the size of the shaded region just above/below each price level, although the height of these regions is no longer linear, but a monotonic non-linear transformation that is visually more illustrative

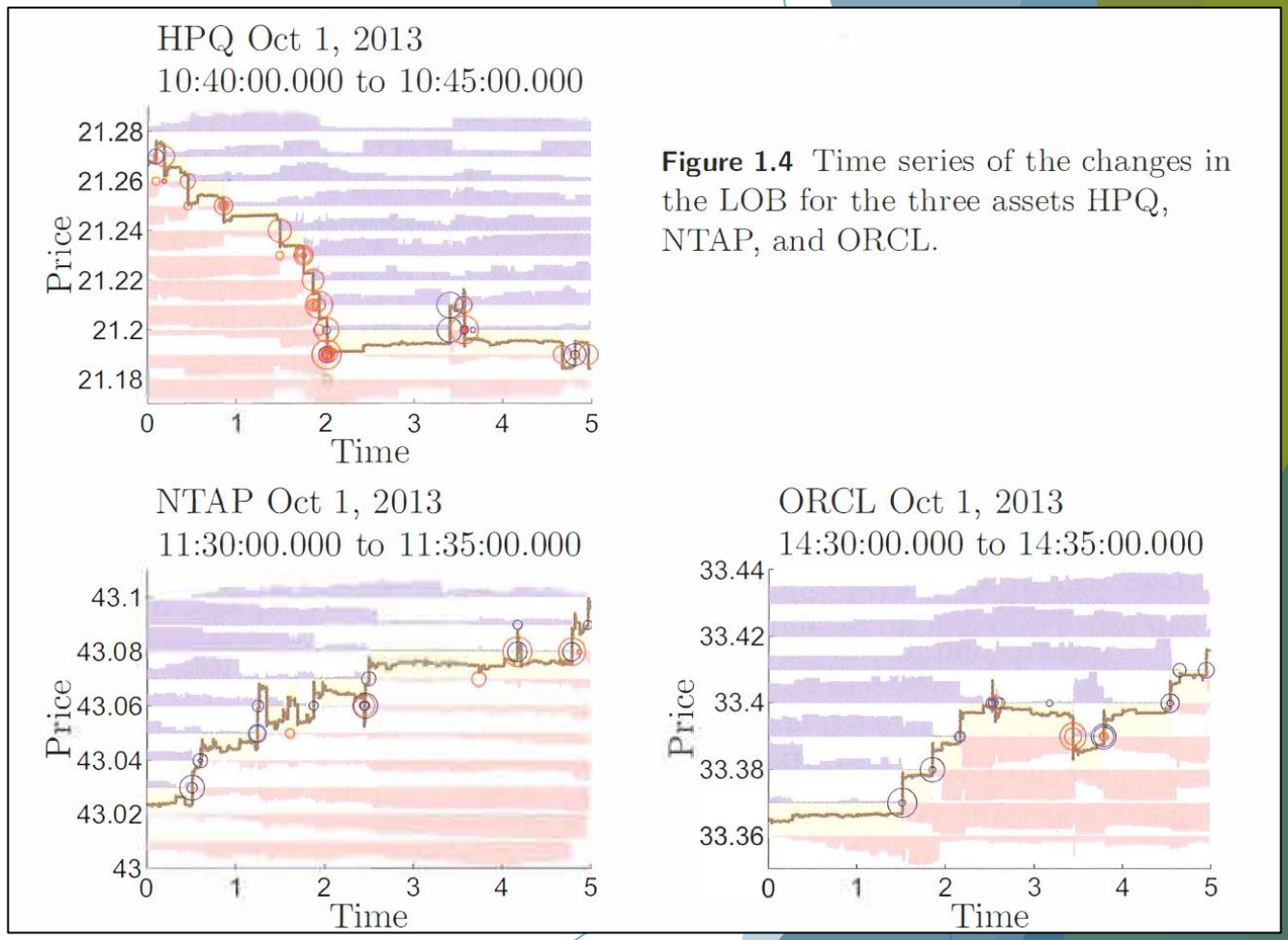


Figure 1.4 Time series of the changes in the LOB for the three assets HPQ, NTAP, and ORCL.

- In addition, the figure identifies when incoming orders were executed.
- The red/blue circles indicate the time, price and size (indicated by the size of the circle) of an aggressive MO which is executed against the LOs sitting in the LOB.
- When a sell MO executes against a buy LO, it is said to hit the bid; analogously, when a buy MO executes against a sell LO, it is said the lift the offer.
- The brown solid line depicts a variation of the asset known as the microprice defined:

$$\text{Microprice}_t = \frac{V_t^b}{V_t^b + V_t^a} P_t^a + \frac{V_t^a}{V_t^b + V_t^a} P_t^b ,$$

- where V_t^a and V_t^b are the volumes posted at the best bid and ask, and P_t^a and P_t^b are the bid and ask prices.

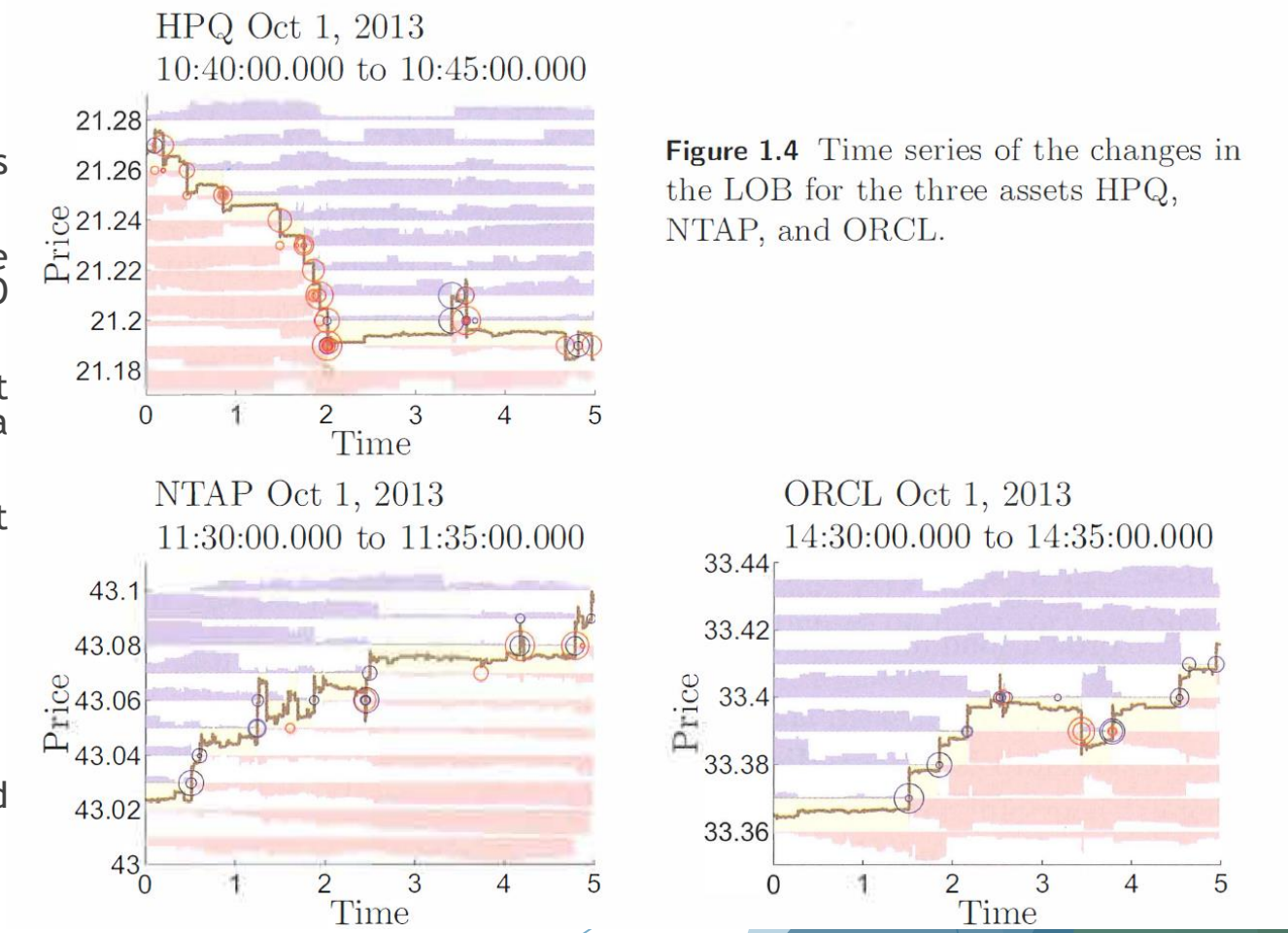


Figure 1.4 Time series of the changes in the LOB for the three assets HPQ, NTAP, and ORCL.

LOB and microprice

- The **microprice** is used as a more subtle proxy for the asset's transaction cost-free price, as it measures the tendency that the price has to move either towards the bid or ask side as captured by number of shares posted, and hence indicates the buy (sell) pressure in the market

- If there are a lot of buyers (sellers), then the microprice is pushed toward the best ask/bid price to reflect the likelihood that prices are going to increase (decrease)

Extended Order Types

- The role of time is fundamental in the usual price-time priority electronic exchange
- Traders need to be able to adjust their trading positions fast in response to or in anticipation of changes in market circumstances, not just at the local exchange but at other markets as well. The race to be the first in or out of a certain position is one of the focal points of the debate on the benefits and costs of 'high-frequency trading'
- The importance of speed permeates the whole process of designing trading algorithms, from the actual code, to the choice of programming language, to the hardware it is implemented on, to the characteristics of the connection to the matching engine, and the way orders are routed within an exchange and between exchanges
- Exchanges, being aware of the importance of speed, have adapted and, amongst other things, moved well beyond the basic two types of orders (MOs and LOs)
- Any trader should be very well-informed regarding all the different order types available at the exchanges, what they are and how they may be used

Some Types of Orders

- **Day Orders**: orders for trading during regular trading with options to extend to pre- or post-market sessions
- **Non-routable**: there are a number of orders that by choice or design avoid the default re-routing to other exchanges, such as 'book only', 'post only', 'midpoint peg', ... ;
- **Pegged, Hide-not-Slide**: orders that move with the midpoint or the national best price;
- **Hidden**: orders that do not display their quantity;
- **Iceberg**: orders that partially display their quantity (some have options so that the visible portion will automatically be replenished when it is depleted by less than one round lot);
- **Immediate-or-Cancel**: orders that execute as much as possible at the best price and the rest are cancelled (such orders are not re-routed to another exchange nor do they walk the book);

Some Types of Orders

- **Fill-or-Kill**: orders sent to be executed at the best price in their entirety or not at all;
- **Good-Till-Time**: orders with a fixed lifetime built into them so that they will be cancelled if not executed by its expiration time;
- **Discretionary**: orders display one price (the limit price) but may be executed at more aggressive (hidden) prices;
- and there are a myriad other variations…
- When coding an algorithm one should be very aware of all the possible types of orders allowed, not just in one exchange, but in all competing exchanges where one's asset of interest is traded. **Being uninformed about the variety of order types can lead to significant losses**. Since some of these order types allow changes and adjustments at the trading engine level, they cannot be beaten in terms of latency by the trader's engine, regardless of how efficiently your algorithms are coded and hard wired

Colocation

- Exchanges also control the amount and degree of granularity of the information you receive (e.g., you can use the consolidated/public feed at a low cost or pay a relatively much larger cost for direct/proprietary feeds from the exchanges)

- They also monetise the need for speed by renting out computer/server space next to their matching engines, a process called **colocation**. Through colocation, exchanges can provide uniform service to trading clients at competitive rates. Having the traders' trading engines at a common location owned by the exchange simplifies the exchange's ability to provide uniform service as it can control the hardware connecting each client to the trading engine, the cable (so all have the same cable of the same length), and the network

- This ensures that all traders in colocation have the same fast access, and are not disadvantaged (at least in terms of exchange-provided hardware). Naturally, this imposes a clear distinction between traders who are colocated and those who are not. Those not colocated will always have a speed disadvantage. It then becomes an issue for regulators who have to ensure that exchanges keep access to colocation sufficiently competitive

Exchange Fees

- Another important issue to be aware of is that trading in an exchange is not free, but the cost is not the same for all traders

- For example, many exchanges run what is referred to as a **maker-taker system** of fees whereby a trader sending an MO (and hence taking liquidity away from the market) pays a trading fee, while a trader whose posted LO is filled by the MO (that is, the LO with which the MO is matched) will a pay much lower trading fee, or even receive a payment (a rebate) from the exchange for providing liquidity (making the market)

- On the other hand, there are markets with an inverted fee schedule, a **taker-maker system** where the fee structure is the reverse: those providing liquidity pay a higher fee than those taking liquidity (who may even get a rebate)

- The issue of exchange fees is quite important as fees distort observed market prices (when you make a transaction the relevant price for you is the net price you pay /receive, which is the published price net of fees)

Three primary classes of traders

- **Fundamental** (or noise or liquidity) traders: those who are driven by economic fundamentals outside the exchange
- **Informed traders**: traders who profit from leveraging information not reflected in market prices by trading assets in anticipation of their appreciation or depreciation
- **Market makers**: professional traders who profit from facilitating exchange in a particular asset and exploit their skills in executing trades

we subsume arbitrageurs into informed traders moving in anticipation of price changes

Classifying market participants

- **Proprietary traders** trade on a (sometimes real, sometimes illusory) trading advantage, which range from the large hedge funds, to small individual 'day-traders' moving in and out of asset positions from their home-offices

- Proprietary traders trade on their competitive advantage: be it identifying fundamentally mispriced assets, identifying price momentum or sentiment-based price changes, having special technical abilities to process market information and identify patterns (technical traders), being able to time price movements based on news (be it the announcement of government economic figures or processing Twitter feeds), or identifying fleeting unjustified price discrepancies between equivalent assets (arbitrageurs)

Classifying market participants

- **Regular investors** and **fundamental traders** are investors who have a direct use for the assets being traded
- They may be individuals who buy stocks in the hope of being able to share in their growth as the corporation increases its economic value-creation and its shares appreciate in value
- Or, they may want to rebalance their investments because of a change in circumstances (in response to a sudden need for cash, a change in their taste for risk or their outlook for the future)
- They may be corporations that use financial contracts to hedge risks such as changes in the prices of inputs and outputs from their production activity

Traders and liquidity

- We can think of market maker types as 'passive' or 'reactive' trading that profits from detailed knowledge of the trading process and adapts to 'the market' as circumstances change
- The other two types represent more 'active' or 'aggressive' trading that only takes place to exploit specific informational advantages gained outside of the trading environment
- A common error is to equate market making with liquidity provision and informed trading with the taking of liquidity
- Market making activity generally favours the provision of liquidity but a particular market making strategy may at times provide liquidity while at others demand it
- Similarly, informed trading does not always occur via aggressive orders, and may at times be better implemented via passive orders that add liquidity

Additional (optional) material: suggested book